Weekly Report

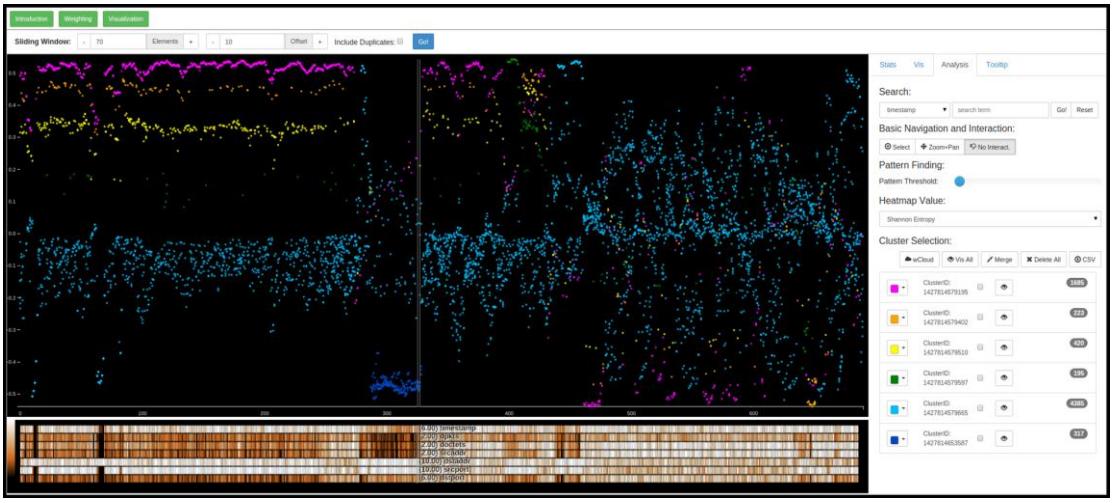
Period: 2016/7/11-2016/7/17
Reporter: Li Zongzhuang

Temporal MDS Plots for Analysis of Multivariate Data **(from TVCG 2016 ,Author** Dominik Jackle and so on)

Temporal Multidimensional Scaling (TMDS), a novel visualization technique that computes temporal one-dimensional MDS plots for multivariate data which evolve over time.

TMDS plots enable visual identification of patterns based on multidimensional similarity of the data evolving over time.

A linked heatmap shows attribute diversity and allows to compare the global MDS patterns for properties of the underlying attributes, supporting the analysis in detail.



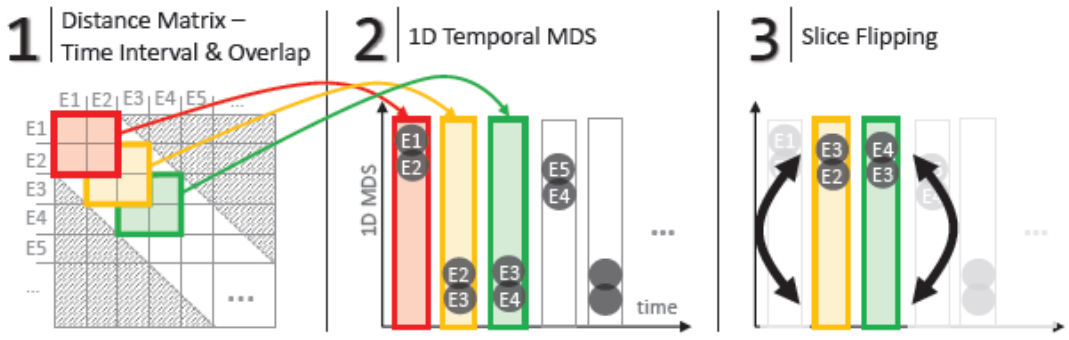


Fig. 2. Consecutive three-step pipeline for the TMDS computation. A

There 's also a case study about network security.

CiteRivers: Visual Analytics of Citation Patterns **(from TVCG 2016, Author** Florian

Heimerl and so on)

The technique uses a highly interactive visualization approach. Through enriching the approach with additional interactive views of other important aspects of the data, the system supports the exploration of the dataset over time and enable users to analyze citation patterns, spot trends, and track long-term developments.

CiteRivers consists of five views of the document set:

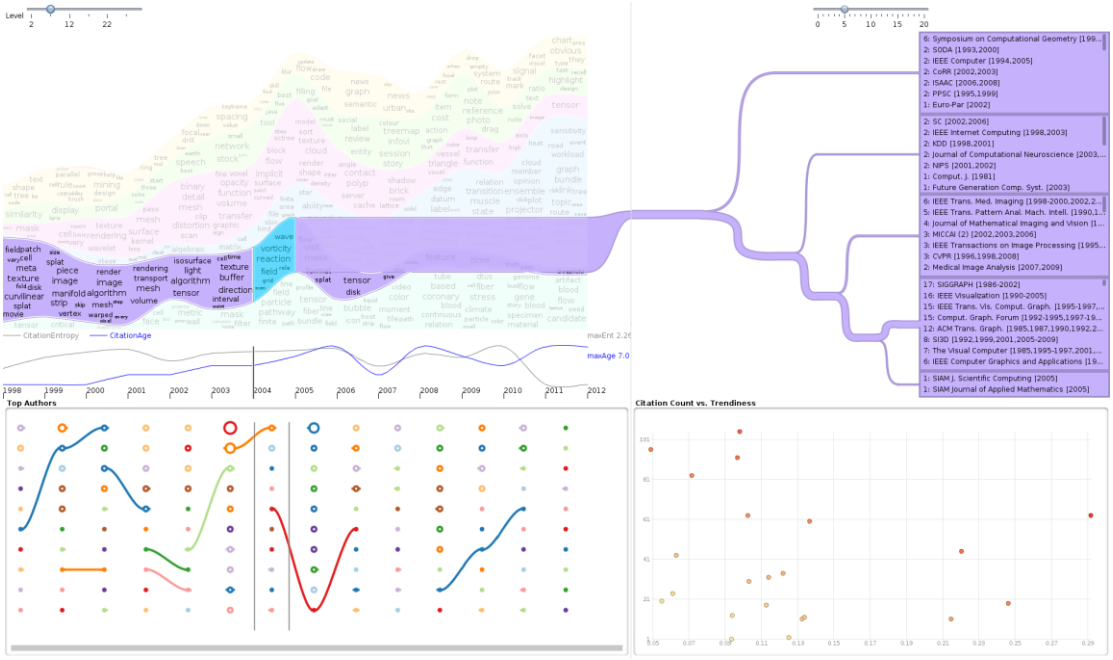
Stream graph Panel

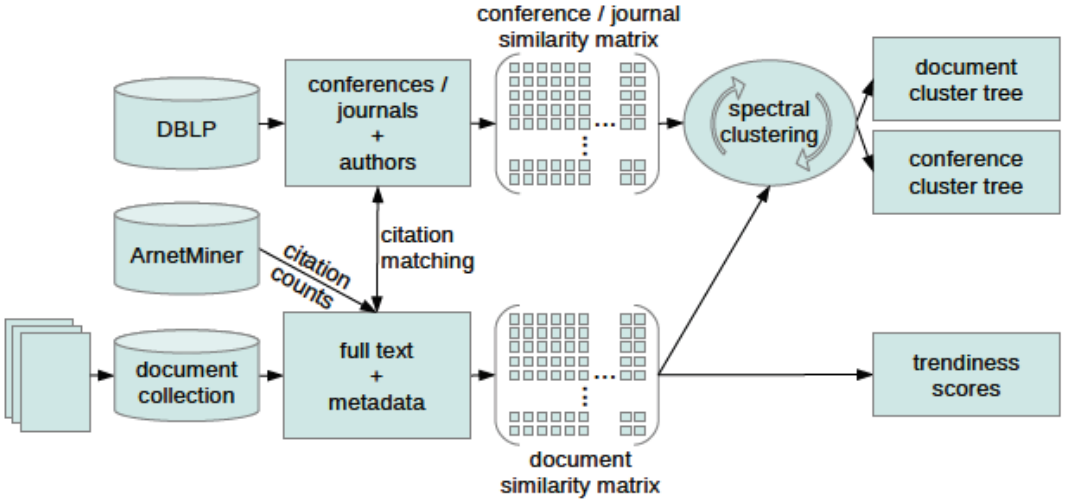
Citation Flow Panel

Citation Aggregation Panel

Author Panel

Document Trend Plot





It also have some data mining methods.

This paper demonstrates the strengths of our approach through a use case and discuss it based on expert user feedback. A dataset of publications from the IEEE VIS / VisWeek conferences covering the years 1998 to 2011 is used as use case.

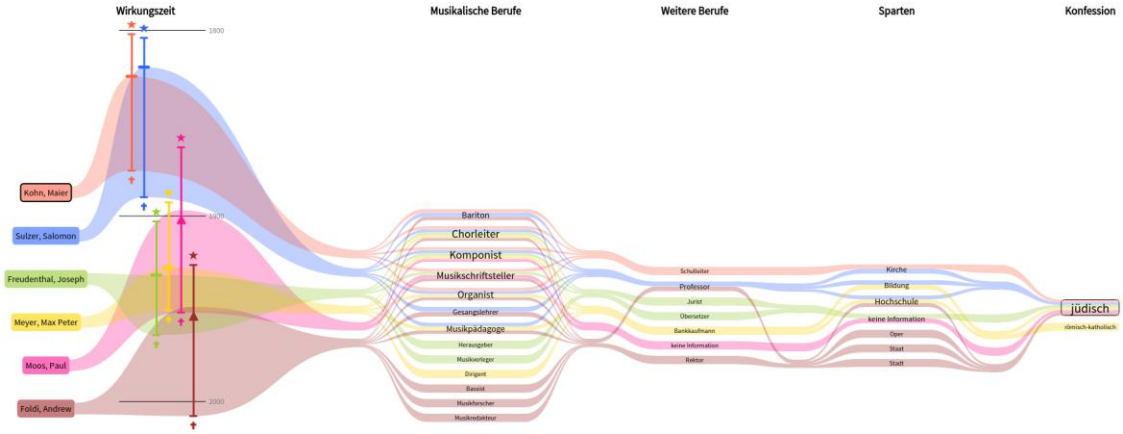
Interactive Visual Profiling of Musicians (from TVCG 2016 ,Author Stefan Janicke and so on)

This paper illustrates the development of a visual analytics profiling system that is used to address the profiling of musicians similar to a musician of interest with the aid of visual means. This is an interesting research question for musicologists working.

Two steps:

(1) the definition of various measures to determine the similarity of musicians' attributes

(2) the design of an interactive profiling system



Contributions:

The similarity of person attributes.

Profiling system

Column Explorer

Temporal uncertainty

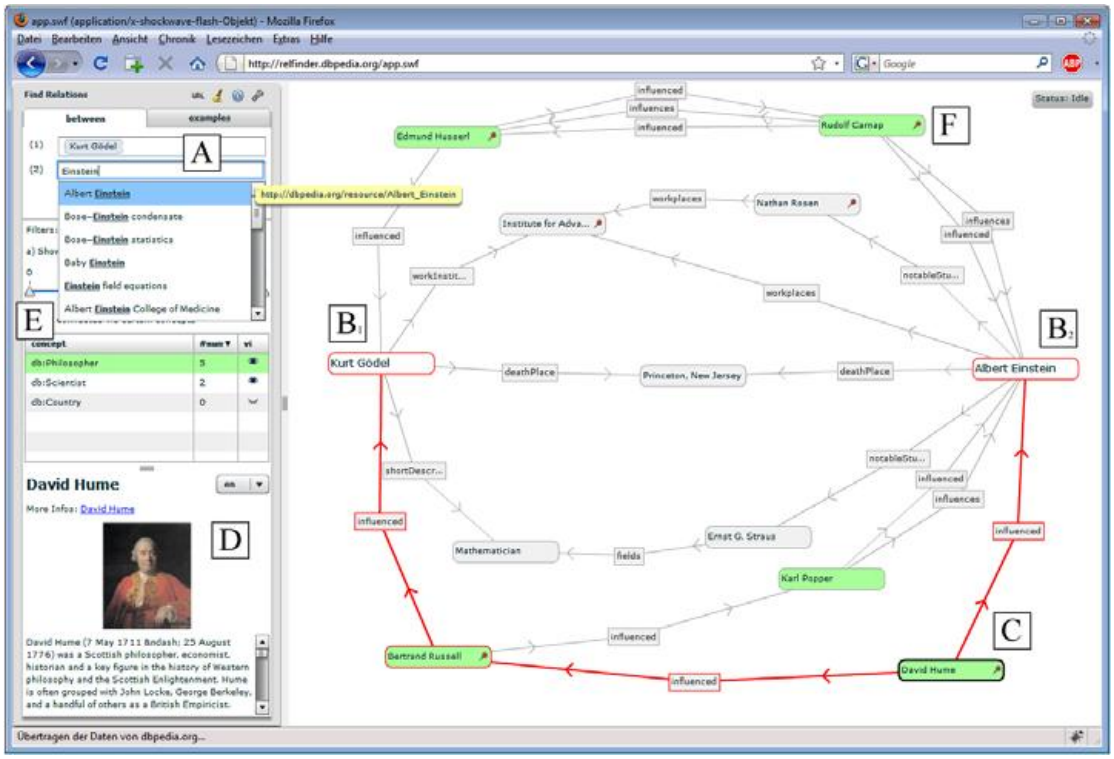
To caculate the similarity of musians, authos use this equation.

$$S(m_i, m_j) = w_p \cdot P(m_j) + \sum_{k=1}^{8} w_k \cdot S_k.$$

And there's different similarities.

RelFinder: Revealing Relationships in RDF Knowledge Bases (Author Philipp Heim and so on)

This paper shows an interactive visualization of this graph that supports the systematic analysis of the found relationships by providing highlighting, previewing, and filtering features.



The basic mechanisms of the RelFinder work with every SPARQL endpoint. It can also be applied to arbitrary knowledge bases.

Linked Data - The Story So Far (**Author** Christian Bizer and so on)

The term Linked Data refers to a set of best practices for publishing and connecting structured data on the Web. Linked Data is simply about using the Web to create typed links between data from different sources.

The concept and technical principles of Linked Data are presented. and these within the broader context of related technological developments are situated.

This paper reviews applications that have been developed to exploit the Web of Data, and map out a research agenda for the Linked Data community as it moves forward.

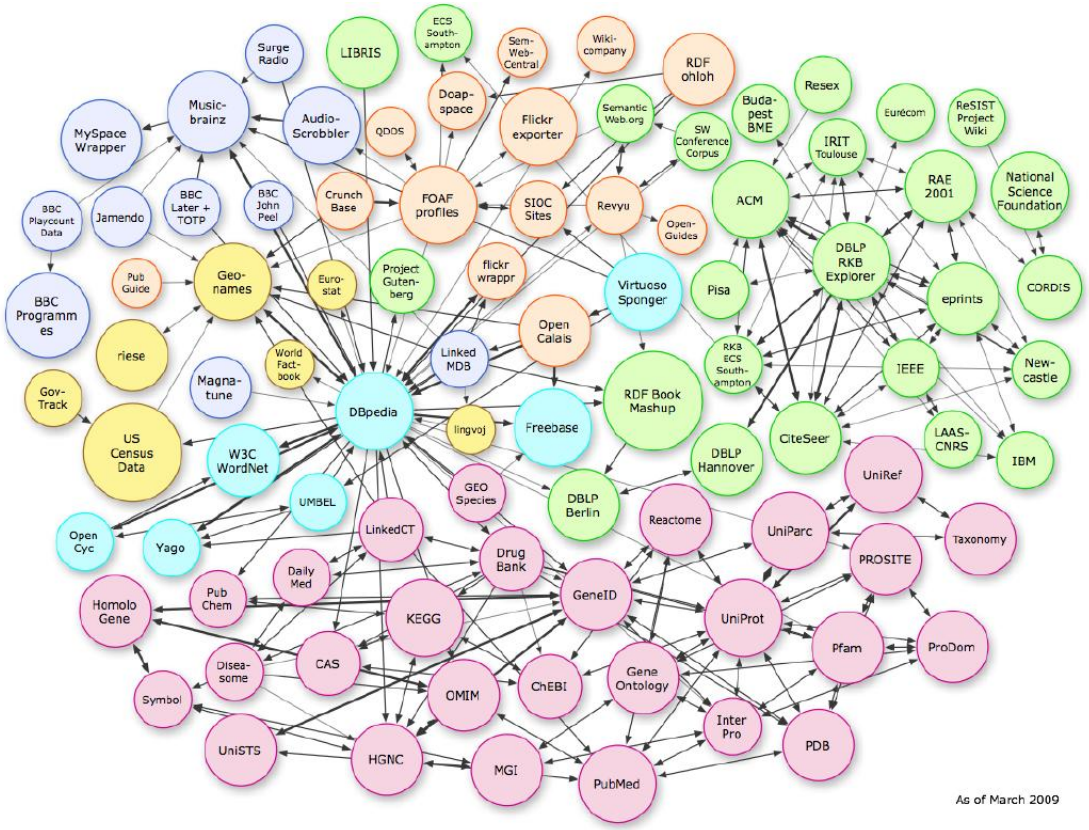


Figure 2. Linking Open Data cloud diagram giving an overview of published data sets and their interlinkage relationships.

Challenges:

User Interfaces and Interaction Paradigms

Application Architectures

Schema Mapping and Data Fusion

Link Maintenance

Licensing

Trust, Quality and Relevance

Privacy

Linked Data realizes the vision of evolving the Web into a global data commons.

附录：

Visual Exploration and Analysis of Knowledge Graph

Li Zongzhuang

Abstract：The Knowledge Graph is a knowledge base used by Google to enhance its search engine's search results with semantic-search information gathered from a wide variety of sources. The Knowledge Graph uses the power of semantics, and wants to improve search precision and effectiveness by building the semantic web. Visualization is the study of (interactive) visual representations of abstract data to reinforce human cognition. The visualization of Knowledge Graph can Effectively

improve the efficiency of the user to complete the search target and precision. Data integration and correlation analysis reasoning is one of the best visual analysis applications. At this time, There are many applications have been exploited based on this concept.

1. Introduction

Knowledge graphs on the Web are a backbone of many information systems that require access to structured knowledge. The idea of feeding intelligent systems and agents with general, formalized knowledge of the world dates back to classic Artificial Intelligence research in the 1980s. Then, with the advent of Linked Open Data sources like DBpedia, and by Google's announcement of the Google Knowledge Graph in 2012, representations of general world knowledge as graphs draw a lot of attention again.

In 2014, Google announced a new initiative,called the Knowledge Vault, which derives much of its data from the Knowledge Graph and the sources thereof, as well as harvesting its own data, ranking its reliability and compiling all results into a database of over 1.6 billion facts collected by machine learning algorithms.

In a paper , author shows what is a knowledge graph:

1. mainly describes real world entities and their interrelations, organized in a graph.
2. defines possible classes and relations of entities in a schema.
3. allows for potentially interrelating arbitrary entities with each other.
4. covers various topical domains.

Vision is the most important channels to the information of the outside world. Visualization is the data technology of interactive visual expression. On the century of big data, The ability of processing data is far behind the ability to get the data. The amount of data contained in Knowledge Graph is huge, so the visualization can be an important means of Knowledge Graph data processing. It can help us find the phenomena and laws faster and achieve the goal. However, the research about the visualization of Knowledge Graph is relatively shallow.

2. Knowledge Graph

2.1 Knowledge Graphs

There are many ways to build knowledge graphs. They can be curated like *Cyc*, edited by the crowd like *Freebase* and *Wikidata*, They can also be extracted from large-scale, semi-structured web knowledge bases such as Wikipedia, *DBpedia* and *YAGO*. Furthermore, information extraction methods for unstructured or semi-structured information are proposed, which lead to knowledge graphs like *NELL*, *PROSPERA*, or *KnowledgeVault*.

Freebase, a public, editable knowledge graph with schema templates for most kinds of possible entities. The last version of Freebase contains roughly 50 million entities and 3 billion facts. Freebase's schema comprises roughly 27,000 entity types and 38,000 relation types.[1] It have been shutdown because of company aquired by Google.

[1] http://www.freebase.com.

DBpedia is a community effort to extract structured information from Wikipedia and to make this information available on the Web. Keys are mapped to properties in that ontology. Based on those mappings, a knowledge graph can be extracted. It contains 6.2 million entities and 187 million statements about those entities. [2]The ontology comprises 735 classes and 2,800 relations.[]

After the shutdown of Freebase, the data contained in Freebase is subsequently moved to Wikidata.[]In Wikidata, for each axiom, it's provenance metadata can be included.[] Wikidata contains roughly 19 million instances and 100 million statements. [3]Its schema defines 23,000 types and 1,600 relations.

Google's Knowledge Graph was introduced to the public in 2012, and it was the tern knowledge graph being coined. Google's Knowledge Graph display was added to Google's search engine in 2012. Once a user search one thing, it provides structured and detailed information about the topic in addition to a list of links to other sites. According to Google, the information in the Knowledge Graph is derived from many sources, including the CIA World Factbook, Wikidata, and Wikipedia.It contains 18 billion statements about 570 million entities, with a schema of 1,500 entity types and 35,000 relation types.[]

Never-Ending Language Learning is an implementation of the Read the semi-structured Web data approach. [] As opposed to DBpedia, all facts recorded by NELL can be tracked according to its provenance and a degree of confidence.[] Nell2RDF platform can transform the data generated by NELL into state of the art Linked Data, following best practices.[] NELL has been learning to read the web 24 hours/day since January 2010, and so far has acquired a knowledge base with over 80 million confidence weighted beliefs (e.g., servedWith(tea, biscuits)).[]

Name	Instances	Facts	Types	Relations
DBpedia (English)	4,806,150	176,043,129	735	2,813
YAGO	4,595,906	25,946,870	488,469	77
Freebase	49,947,845	3,041,722,635	26,507	37,781
Wikidata	15,602,060	65,993,797	23,157	1,673
NELL	2,006,896	432,845	285	425
OpenCyc	118,499	2,413,894	45,153	18,526
Google's Knowledge Graph	570,000,000	18,000,000,000	1,500	35,000
Google's Knowledge Vault	45,000,000	271,000,000	1,100	4,469
Yahoo! Knowledge Graph	3,443,743	1,391,054,990	250	800

An overview about these knowledge graphs.

2.2 Semantic Web

The core concept of Knowledge is the introduction of the semantic, which means that let the computers know the semantic judgments.

3. Visualization

3.1 Graph data visualization

Graph data is an important component in data. The visualizations about graph data are often

[2] http://wiki.dbpedia.org/services-resources/datasets/dataset-2015-10/dataset-2015-10-statistics
[3] https://tools.wmflabs.org/wikidata-todo/stats.php

presented by node-link graph.

3.2 High-dimensional data visualization

There are many data sets have more than one dimension. So many visualization tools have been invented to present high-dimensional data. The results got by Knowledge Graph often have many properties. That means we can get some revelation.

3.3 Another types visualization

Because of the difference of goals, there can be many visualization schemes. Maybe we can learn more from them.

4. Application softwares

Based on the theory of human-computer interaction, there are a lot of software is put forward based on the semantic web. This kind of software focus on data integration and correlation analysis. Data integration made in background automatically, and data correlation analysis mainly rely on people's reasoning ability as well as front end some interactions. That’s the best application, which give full play to the people the calculation of analytical reasoning skills and computer specialty.

There are many applications in this area, such palantir, IBM i2, Tableau and so on.